

**FIG. 1**



Initial strings A, B, and C:

**String A:** A1 - A2 - A3 - A4 - A5  
**String B:** B1 - B2 - B3 - B4 - B5  
**String C:** C1 - C2 - C3 - C4 - C5

↓  
Select substrings

String Pools:

**Pool 1:** A1, B1, C1  
**Pool 2:** A2, B2, C2  
**Pool 3:** A3, B3, C3

↓  
Concatenate  
substrings

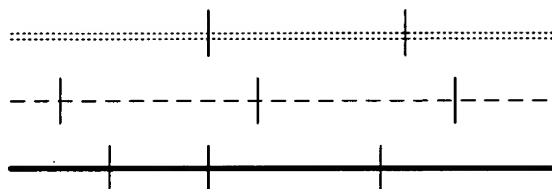
New Strings:

**String A:** A1 - A2 - A3 - A4 - A5  
**String B:** B1 - B2 - B3 - B4 - B5  
**String C:** C1 - C2 - C3 - C4 - C5

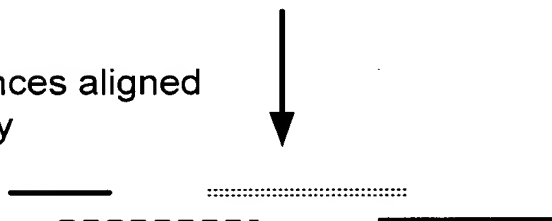
**FIG. 2**



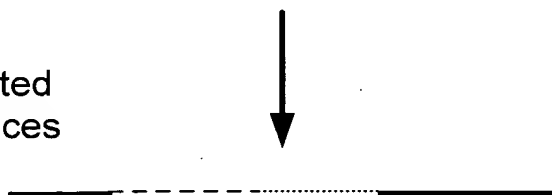
Initial sequences



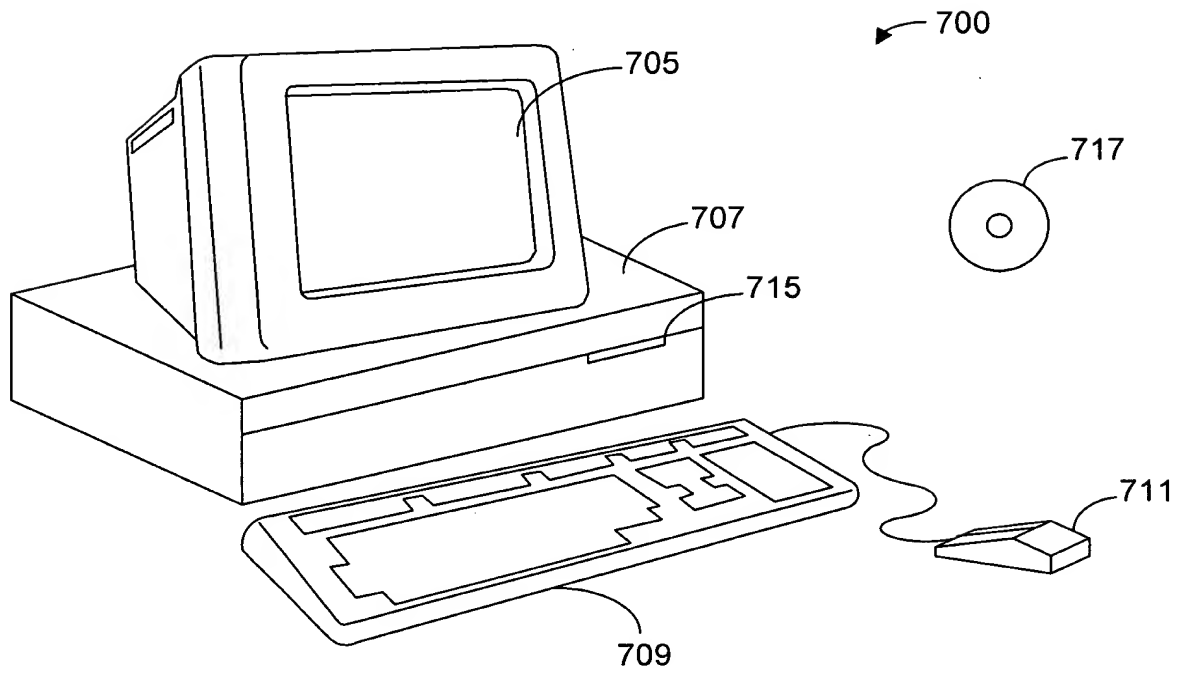
Subsequences aligned  
by similarity



Concatenated  
subsequences



**FIG. 3**



**FIG. 4**



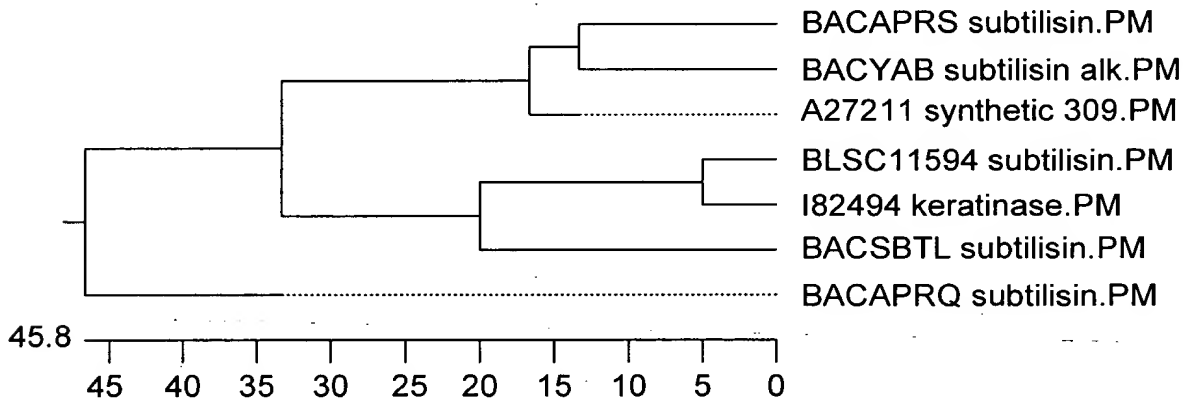
FAMILY GAGGS MODEL # 1.

SUBTILISIN BACKGROUND INFORMATION:

7 PARENTS, SERINE PROTEASES, DIVERSE

TYPE OF ALIGHMENT/SIMILARITY DATA PRESENTED:  
 AMINOACID SEQUENCES, LEADER PEPTIDE EXCLUDED.

Percent Similarity								
	1	2	3	4	5	6	7	
1		62.1	81.4	57.6	81.8	56.1	59.1	1
2	50.5		61.0	54.9	59.5	58.2	60.8	2
3	21.0	52.0		54.6	78.4	50.6	53.2	3
4	54.4	63.3	62.3		52.0	64.6	67.9	4
5	20.5	54.9	25.1	65.6		53.9	56.5	5
6	58.6	56.6	72.2	44.2	63.4		94.9	6
7	52.5	51.4	66.0	38.5	57.8	4.9		7
	1	2	3	4	5	6	7	



**FIG. 5**



FAMILY GAGGS: SUBTILISIN MODEL. PAIRWISE DOT-PLOT ALIGNMENTS TO FIND HOMOLOGY AREAS

LEADER PEPTIDE

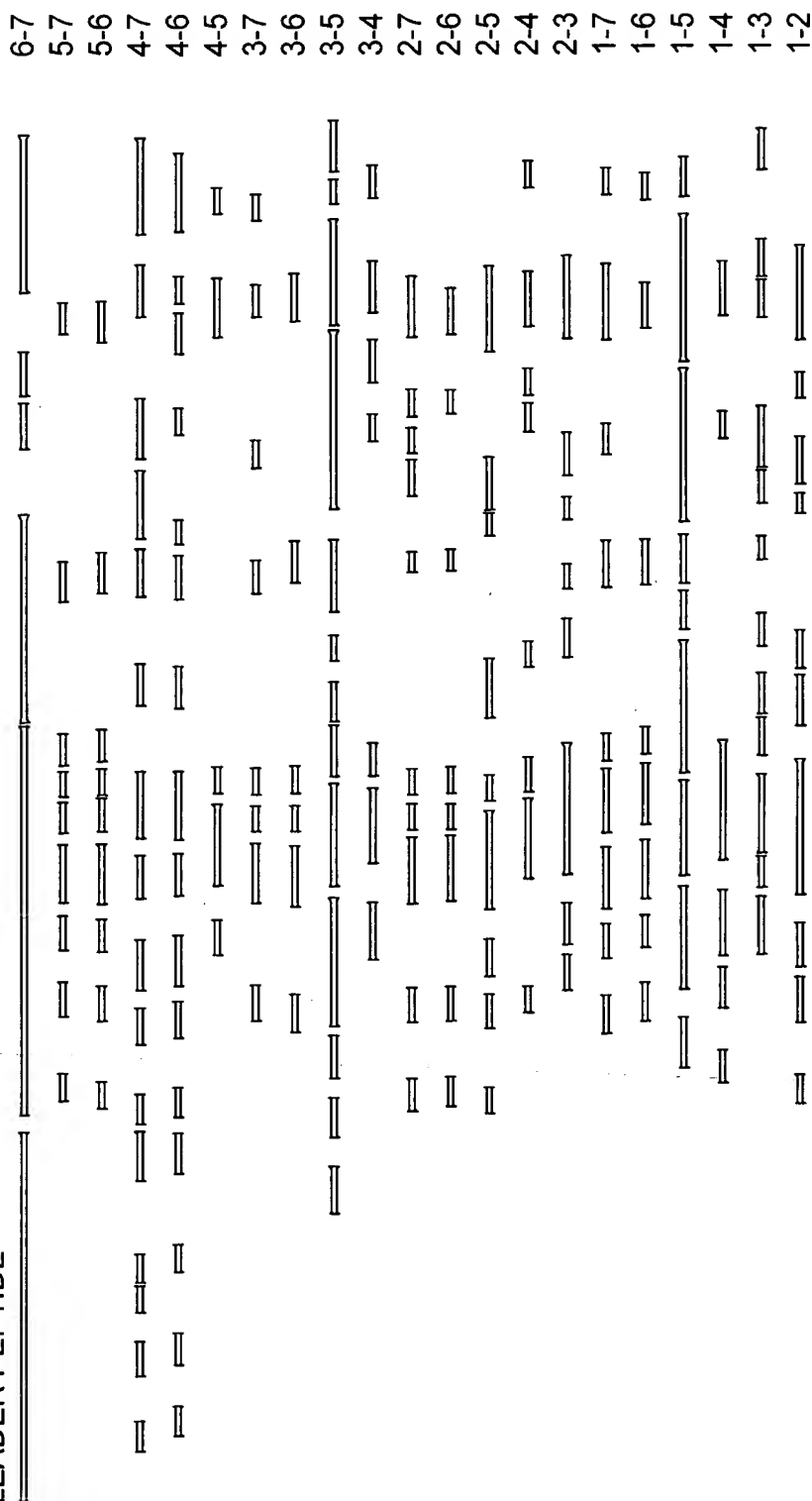


FIG. 6



GAGGS - SUBTILISIN MODEL (7 PARENTS) SELECTING PAIRWISE CROSSOVER POINTS

LEADER PEPTIDE

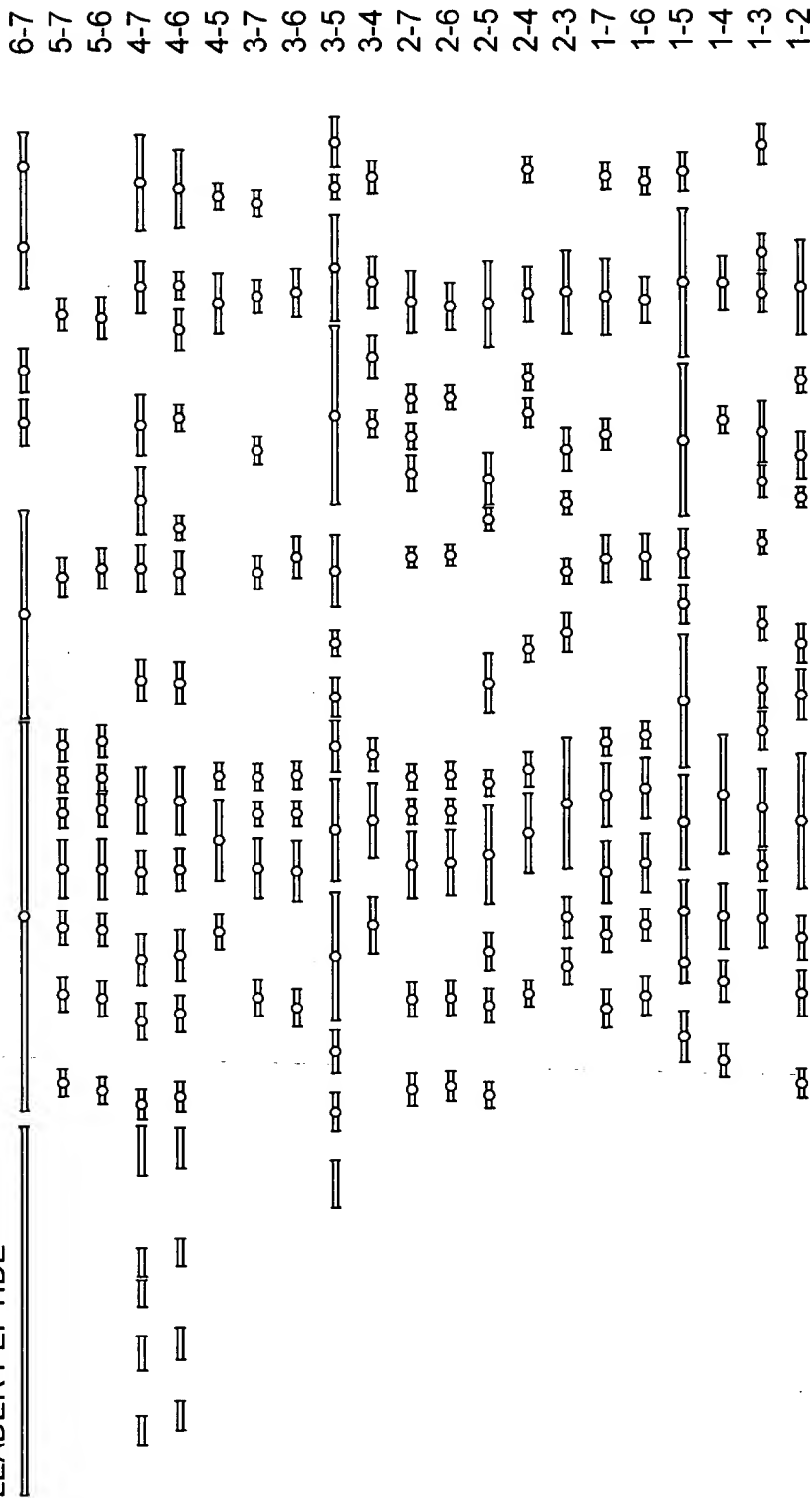


FIG. 7